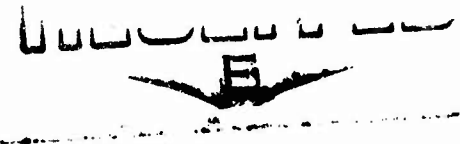AFOSR 67-2809

23

DEC 26 1967

by DR. HAROLD WOOSTER

Director, Information Sciences Directorate,
Air Force Office of Scientific Research,
Office of Aerospace Research

The keynote address
delivered at the sixth annual
INSTITUTE IN TECHNICAL AND INDUSTRIAL
COMMUNICATIONS

COLORADO STATE UNIVERSITY
JULY 1963
FORT COLLINS, COLORADO

49

About Colorado State University's Institute in Technical and Industrial Communications

The Sixth Annual Institute was held during the week of July 8-12, 1963.

The Institute in Technical and Industrial Communications is an intensive one-week course and workshop for writers, editors, journalists, scientists, engineers, administrators, and information specialists, who supervise or prepare technical and industrial communications. The faculty is composed of University staff and, in the main, prominent authorities and practitioners in the field of scientific, technical, and industrial communications.

Proceedings of these annual Institutes can be purchased from the Institute in Technical and Industrial Communications, Colorado State University, Fort Collins.

The Seventh Annual Institute in Technical and Industrial Communications will be held July 6-10, 1964. Program and registration details may be obtained by writing the Director.

A WEB OF WORDS

by DR. HAROLD WOOSTER

Director, Information Sciences Directorate,
Air Force Office of Scientific Research,
Office of Aerospace Research

The keynote address
delivered at the sixth annual
INSTITUTE IN TECHNICAL AND INDUSTRIAL
COMMUNICATIONS

COLORADO STATE UNIVERSITY
JULY 1963
FORT COLLINS, COLORADO

The Frame of Reference

Demonology Discussed

Engineering has been defined as a branch of applied demonology; science as the casting out of demons (1). The great achievements of classical thermodynamics for example may be summarized as the casting out of three demons. The First Law, that you can't get something for nothing, cast out Aladdin's Demon. The Second Law, that even if you do you have to pay for it, cast out Maxwell's Demon—the homunculus who could sort out the hot molecules from the cold without the expenditure of energy. The Third Law, that there is no disorder at absolute zero, cast out LaPlace's Demon—who could predict the entire future of the universe if you could only tell him, for any given instant in time, where every particle in the universe was and how fast it was traveling.

Admittedly, it is not easy to become an applied demonologist. One must spend many years and much of his parent's money in going to places like the Moloch Institute of Tenebriology, where one sits at the feet of learned senior demonologists who teach that by performing specific ritual actions which need not be understood and certainly should never be questioned, one will be able to summon certain inferior demons—superior demons continuing to require the consultant services of a senior demonologist on a standard fee plus per diem plus travel expense basis—from the vasty deeps and, if the book is followed exactly, such demons can be compelled to perform certain specific acts for the benefit of the aforesaid applied demonologist, junior grade.

In this sense, you and I are demonologists. We are word-smiths. Words are our stock in trade. We get paid for our ability to pick words and put them together, trunk to tail, in sentences. Yet, although the Pen-

tagon has replaced the pentagram, and the *Necronimicon* of the mad Arab, Abdul Al-hazred (2) has been replaced by the *New International Dictionary* (3) of the mild lexicographer, Philip Gove, we still do not know what words are, nor sentences, nor, for that matter, what demons we are trying to reach or what we are trying to get them to do. Until such time as we understand the forces with which we deal, we are practicing demonology.

There is a branch of demonology today called *linguistics* which is using the white magic of the large-scale digital computer to harness the black magic of language. Linguistics in the past has been hampered by the all too human inability of linguists to read enough and remember enough of even one language to make statistically valid statements about it. The computer has one major asset—it can read and remember far better than any human or team of humans. It has one major liability which strangely enough is also its greatest asset—it can do only what it is told to do. "You know what I mean don't you?" works with humans. It does not work with computers. The use of computers has forced a precision in language and in the making of statements about language that was previously unknown.

Gulliver the Gullible or Linguistics in Lagado

Certain aspects of this newer linguistics apply to the jobs by which we earn our livings. And, since literature is at least one aspect of linguistics, it seemed only fitting to start with Lemuel Gulliver's Third Voyage wherein he visited the School of Languages of the Grand Academy of Lagado (4). (Those who have not read Gulliver as adults, or who think of it only as an amusing tale of pigmies and giants might better spend their time reading Swift than Wooster.)

-3-

The School of Languages was working even then on problems of words, and problems of sentences. As Gulliver reports:

We next went to the School of Languages, where three Professors sat in Consultation upon improving that of their own Country. The first Project was to shorten Discourse by cutting Polysyllables into one, and leaving out Verbs and Participles; because in Reality all things imaginable are but Nouns.

The other, was a scheme for entirely abolishing all Words whatsoever: nd this was urged as a great Advantage in Point of Health as well as Brevity. For, it is plain, that every Word we speak is in some Degree a Diminution of our Lungs by Corrosion; and consequently contributes to the shortening of our Lives. An expedient was therefore offered, that since W o r d s are only Names for Things, it would be more convenient for all Men to carry about them, such Things as were necessary to express the particular Business they were to discourse on.

And this invention would certainly have taken Place, to the great Ease as well as Health of the Subject, if the Women in Conjunction with the Vulgar and Illiterate had not threatened to raise a Rebellion, unless they might be allowed the Liberty to speak with their Tongues, after the manner of their Forefathers: Such constant irreconcilable Enemies to Science are the common People. However, many of the most Learned and Wise adhere to the new Scheme of expressing themselves by Things; which hath only this Inconvenience attending it; that if a Man's Business be very great, and of various Kinds, he must be obliged in proportion to carry a greater Bundle of Things upon his back, unless he can afford one or two strong Servants to attend him . . .

In yet another room, Gulliver encountered, simultaneously, the problem of sentences, of computers in generative linguistics and the perennial threat of automation.

The next Professor I saw was in a very large Room, with Forty Pupils about him. After Salutation, observing me to look earnestly upon a Frame, which took up the greatest Part of both the Length and Breadth of the Room; he said, perhaps I might wonder to see him employed in such a Project . . . But the World would soon be sensible of its Use-

fulness; and he flattered himself, that a more noble exalted Thought never sprang in any other Man's Head . . . [since] . . . by his Contrivance, the most ignorant Person at a reasonable Charge, and with little bodily Labour, may write Books in Philosophy, Poetry, Politicks, Law, Mathematicks and Theology, without the least Assistance from Genius or Study.

He then led me to the Frame, about the sides whereof all his Pupils stood in ranks. It was about Twenty Foot square, placed in the Middle of the Room. The Superficies was composed of several Bits of Wood, about the Bigness of a Dye, but some larger than others. They were linked together by slender Wires. These Bits of Wood were covered on every Square with Papers pasted on them; and on these Papers were written all the Words of their Language in their several Moods, Tenses, and Declensions, but without any order. The Professor then desired me to observe, for he was going to set his Engine at work. The Pupils at his Command took each of them hold of an Iron Handle, whereof there were Forty fixed round the Edges of the Frame; and giving them a sudden Turn, the whole Disposition of the Words was entirely changed. He then commanded Six and Thirty of the Lads to read the several Lines softly as they appeared upon the Frame; and where they found three or four Words together that might make part of a Sentence, they dictated to the four remaining Boys who were Scribes. This Work was repeated three or four Times, and at every Turn the Engine was so contrived, that the Words shifted into new Places, as the square Bits of Wood moved upside down.

Six Hours a Day the young Students were employed in this Labour; and the Professor showed me several Volumes in large Folio already collected, of broken sentences, which he intended to piece together; and out of those rich Materials to give the World a compleat Body of all Arts and Sciences.

(Although it is not germane to my main line of argument, I must mention here the peculiar fascination that this scheme or its lineal descendants, the chimpanzees with typewriters to type out, eventually, all the books in the British Museum or, presumably, the very large computer with all the words in the English language in its memory, as well

as all the rules for generating English sentences has held for me since I learned that someone, possibly Poincare, showed that by the time they or it had finished writing all the books that had been written, they or it would have written all the books that have never been written!)

As Gulliver said farewell to beautiful Lagado and set sail for Glubbdubdrib, he took with him two linguistic "facts" that he had learned:

(a) Words are the Names of Things.

(b) Sentences are combinations of words that humans recognize as sentences.

How much more do we "really know" than Lemuel Gulliver did?

The Mental Universe

To answer this question, it was necessary first to establish a frame of reference. Each of us, as we crawl like pink four-dimensional worms across three-dimensional space on our way from womb to grave, lives simultaneously in at least three universes: The mental, the physical, and the linguistic.

We can know only one mental universe thoroughly, our own, and we may count ourselves fortunate if we have not had to pay $25 the 50-minute hour to learn that. We can only infer what other mental universes : like by analogy, always a dangerous ول, -how many of the assorted woes of mankind have stemmed for the simple statement: "I think this way; therefore he, or even more dangerously, she must think this way,"—and by observation of response to stimuli. When, in 1913, John Watson closed the "bloodshot inner eye of American psychology" (6), the profession started to train their exteroceptors on the laboratory animal and, eventually on other humans (starting, of course, with graduate stu-

dents.) Thus association psychology became behavior psychology, now testing the response of other organisms on a question-answer, stimulus-response basis.

I am not at all sure that this has helped my personal understanding of my rational and creative processes any more than has the amazing cleverness of the neurophysiologists in poking tiny wires into various nerves and measuring the voltage they find, or the ever increasing feats of "mental" dexterity that computers are being taught to perform.

I do however find it useful to think of the mental universe in computer terms as a "logic" and a "memory," and to borrow from the systems sect that branch of lower demonology I call "black box mysticism"— the notion that if I can describe in fairly precise terms what something does it is not really necessary for me to understand how it does it, unless I want to make another black box to do the same thing (7). The memory then, in my terminology, consists of "tokens," a term deliberately chosen for its ambiguity and lack of connotations—the reader may, as he will, substitute thoughts, ideas, concepts, percepts, images, sensations for "tokens." Some of these tokens correspond, more or less to words, and can be expressed linguistically as words or combinations of words—*Scire est dicere posse*, "to know is to be able to speak;" others most certainly can only be expressed linguistically with a great deal of trouble.

There is a fundamental paradox of mental activity: one must remember in order to survive, one must forget in order to stay sane. Since people do, usually, both survive and stay sane, it is convenient to assume that these tokens are stored in areas of varied functional accessibility—that some are harder to get to than others.

The "logic," corresponding perhaps to

core memory, has at least two different jobs to do. One is just what the name applies—to store the rules for the regular and normal use of tokens. "A implies B" 99 plus percent of the time is the sort of rational operation this logic does more or less well.

Another part of this logic might more properly be called an "illogic." It is, I think, what I mean when I say "I." It is what distinguishes me from you, and both of us from the animals. Its most important role, in the context of this discussion, is simple to state and impossible to define. It recognizes felicity; it tells the writer when he has hit upon just the right combination of words, the artist on the right line, the scientist when he has hit on the right guess for the structure of a complicated organic molecule. (Is it necessary to point out that no two felicity recognizers are alike? Scientists' hunches may be wrong; people do write lousy prose and paint worse pictures.)

There is one other factor recognized by all of us who have ever been stared down by a blank piece of paper. What starts the machinery going? There must be an "idea generator" or a "hypothesis generator" partially, but only partially, under control of the logic which starts the tokens churning in various logical and illogical combinations past the "felicity recognizer." The process is both deterministic and probabilistic, and "I" is not always very good at starting it or turning it off, although it is possible for example for "I" to ask myself, "Where did I see that before?" and forget about the search until the answer pops into conscious attention.

Aldous Huxley has put this far better than I can in his essay, "Knowledge and Understanding," (8), when he writes:

> How do electro-chemical events in my brain turn into the perception of a quartet by Haydn or a thought, let us say, of Joan of Arc? I haven't the faintest idea—nor has

anyone else. Or consider a seemingly much simpler problem. Can I lift my right hand? The answer is, No, I can't. I can only give the order; the actual lifting is done by somebody else. Who? I don't know. How? I don't know . . .

Even my claim to think is only partially justified by the observable facts. Descarte's primal certainty, "I think, therefore I am," turns out on closer examination to be a most dubious proposition. In actual fact, is it "I" who do the thinking? Would it not be truer to say, "Thoughts come into existence, and sometimes I am aware of them?" Language, that treasure house of fossil observations and latent philosophy, suggests that this is in fact what happens. Whenever I find myself thinking more than ordinarily well, I am apt to say, "An idea has occurred to me," or, "It came into my head," or, "I see it clearly." In each case the phrase implies that thoughts have their origin "out there," in something analogous, on the mental level, to the external world . . .

In relation to the subjective "I," most of the mind is out there. My thoughts are a set of mental, but still external, facts. I do not invent my best thoughts; I find them. Total awareness, then, reveals the following facts: that I am profoundly ignorant, that I am impotent to the point of helplessness and that the most valuable elements in my personality are unknown quantities existing "out there" as mental objects more or less completely independent of my control.

This curious collection of goings on in the mental universe is linked to the physical universe by various input and output devices. All information that reaches the mental universe is encoded by the body's own encoders if not before it reaches the body. And remember, a code is a simplified representation of something else, never containing more, and almost always less information than the original.

Output may be either linguistic, paralinguistic, or kinesic. The distinction is a useful one to remember. Consider a sound film of this talk. Take the sound track and give it to a secretary to transcribe. By this definition, the information on the typescript is linguistic; the difference between the infor-

mation on the typescript and the sound track—intonations, stress, pronunciations—is paralinguistic. The information you get from shutting off the sound and watching the silent film of my gestures and body movements is kinesic.

To summarize, then, the mental universe with the possible exception of the one you own and occupy—not as the captain of your soul but only as its noisiest passenger—may be thought of as a sealed box, containing a memory store, composed of tokens and associations between tokens, a logic and a felicity recognizer linked to the store through a random idea generator and connected to the physical universe through a series of encoded inputs and outputs.

The Physical Universe

I intend ⌣ treat the physical universe quite cavalierly. It exists. I take refuge in the naive realism which is said to be the highest form of philosophical sophistication and refuse, at least in this context, to discuss neither whether common sense, in fact, applies to quantum physics; nor the problems of the observer who is himself a part of the universe he is observing; nor whether the object of scientific knowledge can ever be known directly through observation or experimentation or whether it can only be known through speculatively observed theoretic construction or axiomatic postulation, which can be tested directly and experimentally only through its observed consequences (9).

There seems to be no great harm in agreeing with the physicists—it might even make them happy—that all the apparently independent existents in the physical world are built up of a limited number of patterns of identical units of energy, and that an ultimate physical identity underlies all the apparent physical diversity of the physical universe, and that all apparently indepen-

dent existents are, to a greater or lesser extent, interdependent.

It is important to note that the physical universe is even less susceptible to direct observation than the mental universe. We at least have the advantage of living as a privileged observer in one of the many sealed compartments of the mental universe; we are completely shut off from direct observation of the physical universe. We can only form hypotheses in the closed caskets of our skulls, and test them by making pointer readings.

The Linguistic Universe

Error is never so difficult to be destroyed as when it has its roots in language.
Bentham

I have almost completed my metaphor. We have a very large black box indeed, consisting of the physical universe as it is, was, and will be—a world, we hope, without end. We have another very large set of much smaller black boxes, consisting of the private mental universes of those who are, have been and, we hope, will be. One thing is lacking—the set of cables to connect these black boxes into a multi-dimensional network of almost infinite connectivity in space and time. These connections, or more properly the information they convey form the linguistic universe, and, when we come to the linguistic universe, we can heave the same sort of sigh of relief that a repairman does when he finds that Inscrutable Box A is linked to Unfathomable Box B with a multi-stranded cable, all laid out nicely for his test prods. He may not be able to decipher the signals, but at least he has signals to decipher.

Related sorts of signals clump together into languages, and the study of these languages is the sphere of linguistics in the broadest sense. There are all sorts of languages, of course: the natural languages of

human reality; the formal object languages of chemistry and physics; the abstract formal languages of mathematics, which may bypass the physical universe entirely in talking non-sense, but non-sense of the most rigorous sort. There are whole hierarchies of metalanguages—the languages used to talk about languages. All languages, however, possess to a greater or lesser degree the following attributes: a vocabulary of symbols, and a set of rules—call them grammar, syntax, calculus, or what you will—for combining these lesser elements into meaningful statements.

It can not be repeated too strongly or too often that symbols have no meanings in themselves. "Meaning" is a property that a human being confers upon a symbol; it takes two who more or less agree to turn this into useful meaning. A private language is no language at all. A thought expressed in symbols that are the private property of the thinker remains his private property.

Or, as Adam Schaff (10) puts it:

> Every material object, or the property of such object, or a material event, becomes a symbol when in the communication process it serves, within the framework of a language adopted by the persons who communicate, the purpose of conveying certain thoughts concerning reality, that is concerning the external world [*my physical universe—h.w.*] or concerning inner experiences (emotional, aesthetic, volitional, etc.) [*my mental universe—h.w.*] of any of the parties to the communication process.

WORDS AND WORD-BOOKS

Above all else rely on words
Then you can pass through
 the safe gate
Into the temple of all certainty
Where even ideas are wanting.
A Timely word will serve as well.
                              Mephistopheles

The name of a thing is something distinct from the qualities of that thing.
                              Karl Marx (11)

*The sign is fixed arbitrarily, the word
has its own history, owing to which it
lives its own life, independently of us.*
<div align="right">S. L. Rubinstein (12)</div>

Words are the symbols by which we earn
our livings. They are perhaps the largest
and, at least to me, the most interesting
class of symbols. What is a word, and what
is meant by "means" in the phrase, "X
means Y"?

To start off with, neither you nor I
"really know" what a word is. There is a
very precise definition: "A word," some
linguists say, "is a sequence of graphemes
occurring between spaces." This illustrates
the first and second of the very many prob-
lems which one encounters in dealing with
languages. Firstly, that rigorous definitions
are not necessarily useful, and useful defini-
tions are almost never rigorous. Secondly,
that a definition is like a banker who will
not lend you money until you can prove
that you don't need it. Unless you can figure
out that a "grapheme" must be a mark
of some sort, the definition means nothing
to you.

The Definer Defined

Ever since you were very young, you
have been told that if you want to find out
what a word "means" or, more precisely,
how it has been defined by one not too par-
ticular dictionary maker, you should "look
it up in the dictionary." It is a curious fact
that ours, perhaps because of the hyper-
trophy of the English language, is one of
the few cultures where teachers work to in-
still "the dictionary habit," and which in-
cludes definitions of words so common that
no one would conceivably want to look them
up (13).

Dictionaries are, or at least should be,
built on the basis of some specific theory
of meaning. Most defining dictionaries
seem to be based on a model which assumes

a meaning proper ("signification," 'comprehension," "intension" in various terminologies) and the things meant by any sign ("denotation," "reference," "extension," etc.). This dichotomy, which goes back at least to medieval philosophy, was discussed in 19th-century linguistics under the heading of "inner form," and in post-Saussurean linguistics in connection with content substance *versus* content form.

Aristotle Was Class Conscious

Dictionary definitions tend to follow the classical Aristotelian pattern of genus, species, and *differentia*. "A is a member of class B, from which it differs in the following particulars" poses a practical problem which we shall encounter again and again.

"A's" in general may also be members of classes C, D, E and so on; a specific "A thing" at a given time may be a member of any, all, or none of these classes. "Automobiles," in general, belong to the class of self-propelled vehicles. My particular, persnickety Morgan resigns its class membership while I, or preferably my wife, pushes it down the driveway to get it started.

There are other problems with statements of class membership. There is the formal problem of generics; each class in which a concept is placed has its own "family tree." There is the chronic confusion between words and their *nominata*—whatever it is that they are supposed to represent symbolically—so that we are never quite clear as to whether we are dealing with the one or the other.

In spite of all these limitations, the basic logic with which we must deal with words is the classical logic of Aristotle and the Schoolman, the "class calculus" or the logic of terms and classes. A class is defined by its members; in lexicography, it is the class of all words assigned to it by the lexicographer. And, in the logic of classes, all you can

say about a given word is whether it does or does not belong to a given class, or combination of classes: A, A *or* B, A *and* B, or *not* A, *not* B.

Include Me Out

Because of our usual muddle between words and what they stand for, it is important to keep in mind the distinction that Quine made in his *Mathematical Logic* between *inclusion* in a class, and *membership* in a class. Where "x is included in y," it may or may not be the case that a particular "x is a member of y," and where "x is a member of y," it may or may not be the case that that particular "x is included in y."

The "class of cats," for example, is included in the class of animals, but is not a member of it. That is, each particular cat is an animal, but the "class of cats" is not an animal. On the other hand, the "class of Mormons" is a member of the "Class of Christian religions," but is not included in it. It is true that the "class of Mormons" is a Christian religion, but not true that each individual Mormon is a Christian religion.

Dictionary definitions are not necessarily ruled by the problems of formal Aristotelian logic (14). The problem of synonymy might be viewed as a problem in class membership when a word is defined as meaning the same as some other word. I tend to view each word as belonging to a class with one member, itself, and side with Fowler (15) when he writes: "Synonyms, in the narrowest sense, are separate words whose meaning, both denotation and connotation, is so fully identical that one can always be substituted for the other, without change in the effect of the sentence in which it is done. Whether any such perfect synonyms exist is doubtful: *gorse* and *furze* may perhaps be a pair." Fowler goes on to point out that misapprehension of the degree which

words are synonymous is responsible for much bad writing of the less educated kind; that nearly all words are partial synonyms and that synonym books in which differences are analyzed may have been engrossing to the writer but offer the reader nothing but boredom. "Everyone must, for the most part, be his own analyst, and no one who does not expend a good deal of care upon points of synonymy is likely to write well."

Semantic Dictionaries

The problem of synonymy and class membership meet in the so-called "semantic dictionary." Semantic dictionary makers try to combine words into classes, and these, in turn, into even larger classes to come up with a minimum number of very large classes corresponding to some sort of semantic factors. Words are then defined by their membership in the appropriate combinations of these classes. The Western Reserve University Semantic Code Dictionary (16) had, at least as printed in November of 1957, some 214 semantic factors; the current record would seem to be held by R. H. Richen's NUDE (which comes with various Marks) and works with 48 elements (17) and defines money as SIGN: STUFF.

To me, as one who has neither made nor used them, these dictionaries suffer from one major disadvantage. There is no question that those familiar with making them can display fantastic ingenuity in making words fit into their categories. Information is inevitably lost when coding is done, and I am not at all sure that all of the original information can be successfully reconstituted from the code.

On the other hand, since these dictionaries were never intended to be used for the purposes for which you and I use a dictionary—one was built for machine literature searching, the other to provide an interlingua for

mechanical translation—this criticism may not be relevant.

Other sorts of definitions encountered include the *denotative* and *ostentative* methods of pointing at specific things, either described in words (denotata) or in the pretty pictures with which we lure our children into the dictionary habit. ("Marred as it was by the cheap sensationalism of its coloured plates" is my favorite phrase to describe the 14th edition of the *Encyclopedia Britannica* (18).) *rule-giving definitions*: "The rule for 'I' is that it is to be used by each utterer to indicate himself," and various analytic and synthetic methods.

What Your Dictionary Won't Tell You— Behaviorism

There are other theories of meaning which the dictionary makers can afford to ignore, but which we cannot. There is a behaviorist or psycholinguistic school, for example, which admits that although meanings are psychic states completely inaccessible to direct observation, words do affect both actions and emotions. Although, as I pointed out in my discussion of the mental universe, each personal set of tokens and connections between tokens is unique, psychologists have demonstrated the existence of impressive interpersonal norms for the emotive force of many words. Words do have meaning as they affect oneself and as they affect others; a writer ignores this fact at his own peril.

Information Needn't Please

Another aspect of meaning is borrowed from information theory. Information, in short, is a measure of the unpredictable. If you know what an article in, for example, *Reader's Digest*, is going to say, there is no point in your reading it. There is a nasty little game that you can play with any author. Stop at the bottom of any right-hand page not ending with a period and try to guess what the word on the top of the next page

-17-

is going to be. I find that I score particularly well on those books I never finish reading!

Our hypothetical hypothesis generator is certainly at work here: forming hypotheses, on the basis of the previous words read, as to what the next word will be, and checking its guess before moving on to the next word in the string. I suspect that this is also the linguistic basis for good "speed reading;" that this same process of forming hypotheses and checking identities and non-identities is what lets you accept or reject a word at a glance rather than slowing down and reading it letter by letter.

From the point of view of the dictionary-maker, the probability of a term's occurrence is directly proportional to its banality, or meaninglessness. It is possible to analyze other people's writing as alternating pulses of banality, but it's hard to build a dictionary this way.

Linguistic Philosophy and Double Talk

Closely related to this is the view of British linguistic philosophy: "The meaning of a term is its use in the language," or, as Bertrand Russell once said: "Every word is a universal, each use of which is an instance." Words, by this viewpoint, can only be understood in context. This is certainly the way you and I enlarge our vocabularies; we resort to dictionaries only when we can not understand the word in context. This is also a favorite trick of both science fiction writers and double-talk experts: "If you want to fly, all you have to do is aufglabe the freel" is a good and incomprehensible sentence in both of these art-forms.

How Little Is Enough?

This viewpoint both simplif: and complicates matters. We would agree that a word can best be understood in context, and that the size of the context must bear some

relationship to the decrease of ambiguity of meaning. It should be possible then, defining n as the number of words on both sides of a given word to test the question asked by Warren Weaver in 1955: "What minimum value of n will, at least in a tolerable fraction of cases, lead to the correct choice of meaning for the central word?"

This test is, in fact, being conducted every day by those who make and use various forms of KWIC—Key Word in Context—indexes of titles. It poses one small problem for the dictionary maker, however. A basic working vocabulary of say, 10,000 words, would contain 100 million possible pairs of words, if one simplified the computation by considering "AA" to mean "very A" and ignored the fact that AB might not mean the same as BA, a factor which would double this number. The number of possible combinations of 10,000 words taken five at a time, surely not an unreasonable request, amounts to 10 followed by 20 zeros, a number I can not pretend to comprehend.

Let me give you a simpler example, somehow more appropriate to Colorado. How many poker hands can be dealt from 52 cards? (The number is curiously close to R. H. Richen's 48 elements of NUDE, taken 5 at a time.) There are possible (19):

4	Royal flushes
36	Other straight flushes
624	Four of a kind
3,744	Full house
5,108	Flush
10,200	Straight
54,912	Three of a kind
123,552	Two pairs
1,098,240	One pair
1,302,540	Nothing

giving a grand and glorious total of 2,598,960 possible 5 card hands, or, in my analogy, possible 5-term phrases from a vocabulary of 52 words.

I suspect that the size of these numbers shows several things: why people persist in playing poker; why makers of semantic dictionaries or, for that matter, classification systems feel no need to multiply their categories; and, why dictionary makers content themselves with the necessary minimum of quotations to illustrate the use of words in context.

Must We Write Pidgin?

The almost limitless possibilities offered by concatenating a relatively few simple words should not lead us to relax and follow the prescription of the manuals on technical writing and the fog-index boys: "Use short, simple words." There are at least two catches to this—one which I will mention later in my discussion of meaning, and one which is pertinent here. Scientific papers are distributed in two ways—direct from the writer to the reader and via some sort of indexing and reference service. You can reach the readers you know about on direct distribution; you can only reach the readers you don't know about through an indexing system, and, if your paper is not properly indexed, it won't be found.

Indexing is done by people today; it may well be done by computers tomorrow. Both machines and people seem to work the same way in the first level of indexing. They try to pick out those single words in a document which will distinguish it from all other documents. The words used for indexing are, in a statistical sense, the unusual words. There is no point in using "group" for an indexing term unless you can find out whether it refers to a sibling, a chemical, or a mathematical group. "Him fella big box, you fight him he cry," is a perfectly good phrase in pidgin (20); "piano" is probably a better indexing term.

Most of us have been conditioned by Philistines who say that readers will balk if they encounter unfamiliar words. I sub-

mit that if you don't use at least a few unfamiliar words to help the indexers do their jobs, the reader may never find your paper at all.

I have my own hypothesis, completely unverified by any experimental evidence, as to why some readers object to unfamiliar words. It stems from something that C. S. Lewis once wrote about the two ways of reading a poem. In the one way, you use the poem as a stimulus for your own private reveries; in the other, you struggle to understand what the author is trying to say. In Lewis' very British metaphor, in the first case, you get a motor to help you ride your bicycle more easily along old, familiar pathways; in the second, you get someone to guide you along new, unfamiliar pathways. The first obviously involves less effort than the second, and the human animal can be lazy.

Have you ever listened to a conversation in which the operative phrases seem to be: "Djuh get what I mean, hunh?" and a random mix of "Uh-huh" and "Unh-unh," and get the impression that each party was basking in a private glow of understanding the other without, perhaps, doing so at all. I will admit that the paralinguistic and kinesic aids of intonation, stress, gesture and feed-back are probably what help us convey complex ideas with relatively small *speaking* vocabularies, but the writer does not have these crutches.

Writers have t h e i r own version of *Catch-22*. The short, simple words we are told to use let our readers think their own thoughts instead of ours; if we try to make them think our thoughts with precise, and thereby almost by definition, unfamiliar words, our readers may not bother to make the effort.

It would be possible to dismiss this cursory discussion of the various theories of meanings of words with a paraphrase from

Gibbon (21): "Theoreticians find them equally false; linguists equally true; and writers equally useful." There are, however, two more points I wish to make.

The Dictionary Makes You Do The Work

The first of these is a practical matter, which I encourage you to test for yourself. Take a big piece of paper and a good, big dictionary; it works even better with big, bad dictionaries. Look up the definition of a word and write it down at the top of the page. Then look up each of the words used in this definition and write these down on the next line. Take a red pencil and draw lines between these words and the word on the line above. Whenever the same word occurs on two different lines, draw a line between them. Keep on doing this until you get tired or run out of paper. Take a look at what you have.

You will find that you have generated two patterns, rings and trees. A ring is a closed loop, where "a" is defined as "b" is defined as "c" is defined as "a." This is a circular definition of the infuriating sort you find in poor indexes which, under "boats" say "See ships," and under "ships" say "See boats."

Trees are branched structures with the ends connected to nothing else at first glance. In fact, these chains relate to you— to something you must be capable of doing or comprehending.

Let's look at the dictionary definition of "word." "A word is an articulate sound or series of sounds which, through conventional association with some fixed meaning, symbolizes and communicates an idea . . ."

Analytically, each of the functional words in this definition relates to something that *you* do. "Articulate" is an operation performed with *your* vocal organs; "sound" is something heard with *your* ear. I do not

see how the words "conventional," "association," and "fixed" can relate to anything but *your* experience; it seems equally obvious that "meaning," "symbolizes" and "communicates" all relate to *you;* and "idea" must be something going on in *your* head.

In short, all dictionaries are bi-lingual. The normal sort of bi-lingual dictionary translates, say, French to English. What we usually think of as solely a monolingual English dictionary is successful only as far as it can translate English into "you."

Unfortunately, there is little help that dictionaries can give in translating "you" into English, helping you to find the right word precisely to express a particular concept. This is a job for the "hypothesis generator" to do—to f o r m a series o f hypotheses about what it might be called until you hit upon the right one. And heaven help you if you can't spell the word you are trying to look up!

What Words Do

The other matter of which I wish to speak is my concept of what a word does. You will recall my discussion of the mental and physical universes connected by the linguistic universe into an n-dimensional net. Words can serve as tokens within our mental universes and as links between our mental universe and other mental universes and the physical universe. This is not to say that they do the job well. As Aldous Huxley writes (22):

> We are human because we talk, and the universe in which we live is largely a home-made affair, carved out of the given world by our vocabulary and syntax, and re-created by ourselves so as to conform in its structure to the structure of the language in which we happen to have been brought up . . . the well remembered dialect in which we talk to other human beings, the native language . . . in terms of which we do most of our learning, almost all our thinking and even much of our feeling and perceiving. (Our perceiving is hardly ever of events as they are immediately given; it is rather of our own ready-

made, verbalized concepts projected by the perceiver into the outside world and superimposed, so to speak, upon the objects of our immediate experience.) Our dependence on language is such that, for the most of us, words no longer stand for things—rather things stand for words, and objects are treated as so many illustrations of our verbalized abstractions. *No language is completely true to the inner and outer world, to which it is supposed to refer.*

Words Plus Words Make Sentences

By being so long in the lowest form (at Harrow) I gained an immense advantage over the cleverer boys . . . I got into my bones the essential structure of the ordinary British sentence —which is a noble thing.

Winston Spencer Churchill (23)

The Mystique of Sentencehood

A language is more than its vocabulary, no matter how large that vocabulary may be. It must also have a set of rules or usages for putting these symbols together into some sort of larger unit. We call these rules syntax or grammar, and at least some combinations of words put together into large units, sentences.

There is a rigorous definition of a sentence: "A sentence is a string of lexes or word forms having the initial letter of the first word capitalized, and the last letter of the last word followed by a period."

There is an operational test of sentencehood: words put into a sentence are easier to remember than the same set of words in almost any other order. Take, for example, the set of words: "furiously," "smoke," "ideas," "green," "colorless," and "the," and try to remember them. Then write these words down in reverse order: "The colorless green ideas smoke furiously." This is a sentence; its reverse is not. You can remember the one and not the other. There is apparently no requirement that a sentence make sense to be a sentence.

Must One Be Elegant?

There is another curious aspect of sentences in English. Unlike other languages, English is comparatively uninflected. There are relatively few hooks and eyes on the ends of our words to tell us which other words they connect with or modify. We must rely upon the order in which we write words rather than the way we inflect them. Yet in many ways English word order is redundant. As a highly developed and free language, English word order becomes less and less necessary for the conveyance of meaning as contrasted with elegance of expression. We really have no difficulty in understanding the Pennsylvania Dutchman who says, "I will go the house in," nor Winston Churchill's "This is the sort of arrant pedantry up with which I will not put." When young people study misplaced modifiers in a high school grammar course, the recognition of meaning enables them to correct the position of modifiers, *i.e.*, to restore grammatical elegance.

As Mortimer Taube has written (24):

> I think it more true to say that meaning in English is a matter of context as contrasted with order, although any individual context might be ordered in accordance with grammatical rules. The point here is quite important . . . Chomsky and others have argued that all ambiguity of meaning results from ambiguity in grammar. I have argued in *Computers and Common Sense* and elsewhere that sentences may be ambiguous because of an undefined context even though their grammar is correct. The reverse situation is also true. Sentences may have bad grammar without being ambiguous in meaning.

I have heard of a computer program for generating euphonious artificial words to name new drugs; there are certainly many computer programs for generating grammatically correct English sentences; I know of no computer programs for generating or recognizing felicitous sentences, although I will admit that the Librascope Auto-Beat-

nik does a surprisingly good job in writing dada or surrealistic poetry. This is not to decry the importance of "good" writing. There is too much of the other kind. It is only to admit that I know of no operational definitions for it.

It's Art, but Is It Pretty?

The "Felicity recognizer" in my hypothetical mental universe knows when it has hit upon a good thing; I do not know how it knows. There are, however, two ways in which you can test how well it is working.

I borrow the first from Robert Graves, who tells somewhere of how he studies the writings of other poets by copying them out in longhand. He claims that this is an almost infallible guide to spotting where the other poet has been trapped into forcing a rhyme, or tried to disguise the fact that he was stuck for the right word or phrase.

Most of us are not poets, nor do we have the time to copy our writings out in longhand, nor in the real world for that matter, do we have the time to follow the excellent and neglected device of putting a piece of work away for a month or so and coming to it as a critical stranger.

Those of us who compose on the typewriter would certainly benefit if we did not have secretaries, and had to re-type what we wrote. You can cut out a lot of superfluous words if you have to do the typing yourself. The easy writing of dictating can make for especially bad reading.

There is, however, one simple practical thing to do with any piece of writing after you have finished it. Read it aloud to somebody. Both the verb and the predicate are important; it must be read aloud, and there must be someone listening. (This is probably why the Ancient Mariner made such

a nuisance of himself at the wedding—he was looking for audience feed-back on a random sampling basis.)

Reading a piece of prose aloud works, and I think I know why it does. One obviously must do it for a speech. Each of us has his personal cadences of intonation and stress in spoken English; one's sentences must match these speech rhythms, or much of the paralinguistic information is lost. I do not envy the ghost-writer who must write a speech for someone else to deliver nor, for that matter, one who has to deliver a speech he has not written. Is it tautological to say: "Good English can be read aloud?" "You write just like you talk" is not a pejorative statement.

The reason why writing that may never be read aloud in public should a l w a y s be read aloud in private is a little more esoteric. Writing tends to be done in fits and starts, in bits and pieces. Speech is continuous discourse. I suspect that running the angular, multijointed monster that comes out of your typewriter past your vocal cords—the "felicity recognizer" working all the while—is perhaps the best way of grinding off the welds and producing, if not a micro-finished, at least a tolerably acceptable piece of prose.

Enough of the quagmire of aesthetics and the problems of producing practical prose. Let me ascend into the empyrean of the theoretical and then come back down to the real world again by discussing, firstly, syntax and secondly, certain formal definitions of sentences.

Syntax Is Taxing

Syntax, speaking very loosely indeed, is the body of rules by which words are put together to form sentences. (The discussion which follows is taken in large part from Reference 17.)

There are at least three operationally differing conceptions of "syntax."

Syntax as Opposed to "Grammar"

This is the old-fashioned grammarian's definition, according to which a *grammatical* operation in any language (such as declension, conjugation, or concord) operates within a word, whereas a *syntactic* operation (such as the government of a noun by a preposition or a verb in a subordinating clause by the subordinating conjunction) operates between words and within a sentence.

This distinction ceases to be important in the computer processing of natural language text, as is done in mechanical translation, since one of the first things that computer programs for analyzing languages tend to do is to split up the natural words you and I know into smaller pieces called morphemes or chunks, since there are a lot fewer kinds of pieces to work with than there are combinations of these pieces into words. It is easier to store the pieces and the rules for making words out of them than it is to store complete words.

When you do this, you find that you have two different sorts of chunks or parts of speech:

a. A large class of *stems*, or *root-words*, or *content-words*, or whatever you like to call them.

b. A small class of *auxiliaries*, or *particles*, or "bits-and-pieces" of language, which can be hooked up to the stem words to make natural language words.

And, you find that the grammarian's definition of grammar as "operation within a word" disappears as you field-strip words into their components.

Syntax as Opposed to "Semantics"

In one sense, syntax (including grammar) is what can be formally handled in a language; semantics is what cannot.

-28-

Alternatively, syntax is the set of all these features of a language which can be handled without reference to "meaning," semantics is what is left over.

In both of these definitions semantics is always negatively defined. They do not talk about what semantics is, only about what semantics is not.

Syntax as Formal Logic

Syntax can be viewed logically, and therefore generally, as the name of a set of logical operations which can be performed on Language, as in the title of Carnap's book, *Logical Syntax of Language*.

In this logician's sense of "syntax," syntax is defined as some sort of permitted operations performed on the units in any formal system which can be interpreted as a language.

In this sense, the syntactic operations can always be precisely defined. What cannot be precisely defined is the notion of interpreting the syntacticized system *as a "language."*

In other words, formal syntactic analysis in this sense in a language-system designed to be interpreted as a real language is a formal set of operations performed upon the units of a system and nothing more. Bar-Hillel has shown conclusively that as soon as you are dealing in any genuine sense with "real language" you need extra-syntactic as well as syntactic information to enable a machine to detect and to distinguish sentences as opposed merely to well-formed formulae. A computer cannot deal with "You know what a sentence is, don't you?" and even as in the School of Languages of the Grand Academy of Lagado, "sentence-hood" may still be something that it takes a human to recognize.

Words and Syntax Equal
Sentences—Sometimes

Notwithstanding the highly subjective nature of "sentencehood," let me try my hand at three formal definitions of sentences:

1. A sentence is a statement in the propositional calculus.

2. A sentence is a basic semantic progression of topics and qualifiers of these topics.

3. A sentence is a discrete fragment of a correlational network of "thoughts," expressed as a sequential linear array of symbols.

These definitions as stated are essentially my own. The very first step needed to bring them into accord with the rigors of their original authors would be to make sure that "is" is always read as "may be regarded as," and *never, never, never* as "is nothing but." The second is to recognize that these statements may be only statistically valid; and all or none of them may apply to any given sentence. With these *caveats* nailed to the masthead, let us proceed.

A Sentence Is a Statement in the Propositional Calculus (25)

In the discussion on words it was shown that the basic logic with which we must deal with individual words is the "class calculus," or the logic of terms and classes. The interest in modern logic, particularly symbolic logic, has shifted to a concern with propositions in which terms are combined into propositions or statements. It is at least tempting to reason by analogy and say that this difference between the class calculus and the propositional calculus must resemble the difference between words and sentences.

For example, a central grammatical conception in all language, natural and artificial, is that of applying a predicate to a subject or subjects; that is, of "saying some-

thing about something." Its clearest case is that of ascribing some property to a single named subject: "Martin is hungry." Although different languages handle the situation in different ways, the situation need not be much more complex syntactically when the predicate asserts a *relation* between two objects: "Martin ate Pamela's cooky." All that is required is to specify the relation and to provide a way of distinguishing the various participants in the tableau.

Inflected languages use a system of word endings; English uses isolating articles and prepositions; symbolic logic must use order alone. It should be possible, with some ingenuity, to map the various predicational patterns of English; verbal, prepositional, adjectival and substantive upon the bare n-adic predicates of symbolic logic, except for one minor practical problem.

One gets the impression that the symbolic logician dealing with natural languages is like the physicist asked to play croquet who says: "Really, I'd much rather play roque"—roque being an advanced form of croquet which, even though the wickets are only 1/8th of an inch larger than the balls, and all sorts of bank shots can be made off the 3-inch concrete curb surrounding the court, is at least played on a dead level surface with precision ground balls so that all shots follow the rules of elastic collisions—and finds himself involved in a particularly savage form of the game described in "Alice in Wonderland," with the hedgehogs used as subjects and predicates, when they are not being violently antagonistic, or quite the converse, turning around and nipping the players on the ankles, and the flamingoes, used as verbs, suffering from acute protein deficiency and not really caring where they get their essential amino acids. "Would that croquet were ruly" is a characteristic utterance in such circumstances.

Notwithstanding these vocational haz-

ards, brave souls can be found to carry on the good work. Bohnert of IBM is trying to work from symbolic logic towards natural language; Walter Sillars of the National Bureau of Standards is trying to reduce English to logical formalisms. There is the Chomsky hypothesis that a full language can be mechanically constructed by deriving it mathematically from a small set of key-forms or kernels; Zellig Harris and his co-workers at the University of Pennsylvania are busily trying to develop rules for reducing complex English sentences to such kernels by a process not dissimilar to the "parsing" we were taught in grammar school.

Logic, as someone has said, can only prove that statements are false, or at least contradictory; it can not prove that they are "true" and stay within the logical system.

There is no doubt in my mind that this approach, the reduction of sentences to statements in the propositional calculus, can teach us the skeletal structure of a sentence and perhaps reveal certain logical fallacies. The well-defined formulae it will someday yield must still be fleshed out with extra-syntactic information to give us sentences we can use.

A Sentence Is a Basic Semantic Progression of Topics and Qualifiers of these Topics

Take care of the sense, and the sounds will take care of themselves.

The Duchess

The precise converse of this statement is the basis of the so-called Guberina hypothesis or "semantic square" (17); so called because I am not at all sure that the inventor would recognize it in its present rapidly-evolving form. Pietar Guberina is Director of Phonetics, Zagreb, Yugoslavia, where he specializes in teaching deaf children to speak languages other than their own. Gu-

berina believes that each language has a characteristic pattern of intonation which it uses to convey meaning, and that the basic overall intonational rhythms of any language are the first and easiest thing for a foreigner learning the language to pick up and understand.

Spoken English, for example, has two "tunes," falling and rising. The falling tune is declarative, complete and confident; the rising tune is hesitant, unfinished and doubtful. These "tunes," unfortunately, do not correspond to known grammatical structures.

Guberina's hypothesis is that at the basis of all human communication there is a single semantic pattern consisting of four inferentially and sequentially connected points which he calls the "Semantic Square." This pattern is prior to all grammar and syntax and is what the hearer—more especially the deaf hearer—tries to get hold of in order to try to understand what is said to him. Guberina believes that from the combination and variation of such squares, all human communication as found in spoken language can be built up.

The easy case to find of a semantic square is that of two paired predicates:

<p align="center">Ann cooked supper;
but
Pamela fixed breakfast.</p>

There are four primary points of semantic intersection:

(1)	(2)
Ann	
Proper Name	supper
GIRL:	FOOD:
NAME	
(3)	
Pamela	(4)
Proper Name	breakfast
GIRL:	FOOD:
NAME	

These four are Guberina's "square points." They are also the points which carry the primary stresses when the sentence is pronounced in English. It can be argued that the full stress-points or choice points of an utterance give the semantically important words, the essential message, and it may well be possible that the essential words of a message form a noisy semantic system—the elements of which have to remove the noise from those preceding them.

The Guberina hypothesis is, at the very least, a fresh semantic basis for human communication. If the primary task of the receiver of any communication is to "Find the Semantic Square" and if the text gives him sufficient phonetic, syntactic, and semantic indications, he, himself, can fill in semantically the remainder of the pattern. He only needs to know the characteristics of squares, not the facts of all human knowledge, to enable him to do so. And, if he can once sufficiently detect the presence of an incomplet· semantic square, he can complete it by adding new semantic intersections or antitheses which, until that communication was made, never existed in any language. He can cause the language to evolve. He can say something new. And no other hypothesis of the nature of human communication meets this difficulty, even in principle.

This is not to say that all sentences should consist of sequences of semantically paired predicates; if they did, it would either look incantational, like the Hebrew Psalms, or nightmarish, like wading through endless reaches of Longfellow's *Hiawatha.*

The Guberina hypothesis, if proven, has several other possibilities; for example: that the phonetics of intonational form gives a theory of meaning, not of sound; that it may model the basic language as given by stressed speech, not the far more complicated forms of language, stressed and unstressed, that we produce when we put

words down in writing. To be dramatic, it may even have uncovered the ancestral ape, the *ur*-sentence, that lurks beneath our civilized prose.

A Sentence Is a Discrete Fragment of a Correlational Network of "Thoughts" Expressed as a Sequential Linear Array of Symbols

This is an eclectic definition, to say the least. It is my own attempt to put together various concepts, borrowed from the Italian Operational School of Language Philosophy (26): the way a Turing machine works, certain characteristics common to most methods of mechanical translation, the properties of the speech signal, and the practical engineering of consoles for command and control systems. It subsumes the two previous definitions of sentences as special limited aspects of this more general definition.

Let me explain what I mean.

A Sentence Is a Discrete Fragment . . .

This rather begs the question. We know that written English is built up by combining individual letters into words and separating these words by spaces, and that at various intervals some of these words are capitalized and that others, usually but not always uncapitalized, are followed by periods, question marks, or exclamation points; and that people can recognize some of these combinations of capitalization and punctuation as sentences, although they have a difficult time telling you how they do it.

Notwithstanding this vagueness, it seems important to include a statement to the effect that a sentence is usually longer than a word, somewhat smaller than one's entire output of written language, and probably somewhat closer to the former than to the latter in size.

A sentence is a piece of something bigger. A piece of what?

... A Correlational Network
of "Thoughts" ...

The Italian Operational Approach to Mechanical Translation is based on the assumptions, as stated in references 26 and 27, that language is an expression of thought and trains of thought, and that these trains of thought are linked together into a correlational network.

Parker-Rhodes of the Cambridge Language Research Unit has shown that certain basic elements of language can be treated as points on a finite lattice, hinting strongly that language is a thesaurus-web.

Psychologists do seem to enjoy inventing and running word association tests even though I am not at all sure what these measure beyond the size of the test subject's vocabulary and his response time.

All of us do verbalize concepts, and conceptualize words.

One would seem safe in assuming then that language does form a rather intricately tangled network of "X's;" "thoughts" would seem to be as good a word as any to substitute for "X," but the reader should feel free to substitute any word he happens to prefer.

... Expressed as a Sequential Linear ...

"Sequential" and "linear" are the operative words. Their order is unimportant. What is important is that they state the single most important limiting factor in the expression of language. However rich the correlational network from which it springs, however luxuriant the associations the reader brings to it, language is put down one word at a time, trunk to tail, and read one word at a time, trunk to tail. (Even in speed reading, where great gouts of text are blurred. and perhaps even the peripheries of the lines omitted, one still reads the text in the linear order in which it was set down.)

As Kipling wrote, of quite another matter entirely (28):

"For the race is run by one and one, and never by two and two."

I am not enough of a topologist to know if it is theoretically possible to compress an n-dimensional lattice or selected portions thereof into an 1-dimensional linear structure to pass through the needle's eye, separating writer from reader, speaker from hearer, one word at a time and then a re-expand it into a reasonable facsimile of the original. It may well be impossible, yet it is what we must do when we communicate our thoughts to others by language.

... Array of Symbols ...

Array is the operative word. It is used here to refer to the possibility of simultaneously displaying two or more symbols relating to the same event. Arrays offer a partial solution to the linear compression problem above.

The speech signal is a particularly rich form of array for displaying information about the message and about the speaker. The bare linguistic bones of the speech message can be conveyed by the transcript; the paralinguistic difference between the spoken and written word can both emphasize the points which the speaker thinks are important, but can also tell the trained observer a great deal about the physical and emotional state of the speaker, as well as his racial, regional and educational background.

The printed word has nowhere near these resources. There was, to be sure, a fad in the seventeenth century for making verses assume grotesque shapes and sizes having something to do with the theme of the poem: hearts for love songs, wineglasses, bottles and casks for drinking songs; pulpits, altars and monuments for religious verses and epitaphs (30). Perhaps the best known of these is the Tale of the Mouse, in *Alice in Wonderland*, a poem of 52 lines set serpentinely in type sizes starting with 12

pt., and with the very end of the tale in 4 pt. A certain amount of respectability is lent to the field by the fact that the logician and philosopher, Charles Peirce, was much interested in the visual analog of poetic onomatopoeia. Among his unpublished papers there is a copy of Poe's *The Raven*, written with a technique that Peirce called "art chirography." The words were formed so as to convey a visual impression of the poem's ideas (31).

Today, however, one is most likely to encounter recognition of the fact that type can be used to convey parallel messages in children's books: "Martin heard a LOUD NOISE." The IBM mechanical translation system prints the transliterations of Russian words it can not translate in red, and stacks synonyms for words it is not sure of one over the other, so that the reader can make his own choice from context.

The concept of arrays of signals relating to a single symbol is important to remember for the future—even if only to torture computer salesmen—although it is probably not worth the trouble for most practical prose.

Sentences Plus Sentences Make Prose

It is difficult enough, oneself, to survey this whole, whether nature or art, but still more difficult to help another to such a comprehensive view. This is due to the consecutive nature of the only methods available to us for conveying a clear three-dimensional concept of an image in space, and results from deficiencies of a temporal nature in the printed word.

Paul Klee

The conscientious writer must not only be able to take a complex mental image and squeeze it down into a single string of words, hoping against hope that the reader will be able to expand this surrogate into something

resembling the original image, but he also has an even more difficult problem.

One would at present have some difficulty in designing a black box to bridge the gap between a 3-megacycle television coaxial cable and a 3-kilocycle teletype line, so that the literate reader could learn as much about a show from the typed output as the illiterate viewer could from his monocular hypnogogue.

Yet this problem is trivial in comparison to the problem of the writer who must use this highly limited tool of linear prose to describe a complex pattern of events occurring at widely separated points and times or, perhaps even worse, more or less simultaneously, by trying to braid and again compress these linear strings of prose into one straight line of expression.

The *Times Literary Supplement* concerned itself, *inter alia*, with this problem in a series of seven articles on various aspects of the interactions between computers and the printed word during March-June 1962 (33). One of these articles "By Our Special Correspondent" dealt in particular with the problems of "Poetry, Prose and the Machine"—my own title for it might have been "Beauty and the Electronic Beast." Quotation marks should be assumed around much of what follows:

An art historian, Herr Werner Haftmann, has apologized thus for the shortcomings of our prose technique:

> It is solely his medium—logical discourse—that compels the historian to follow the highly unpleasant "linear" method. He proceeds from stone to stone, and for the sake of deductive clarity easily leads his reader to mistake the abstract "line" for the whole of reality . . . To write from a simultaneous point of view, that is to say, to grasp reality in its complex simultaneity, is technically almost impossible for an historian who has to deal with a vast amount of material.

Here is the real problem for the writer. Tradition, abetted by humanist education, has led us to think of narrative prose as the best instrument for conveying an intricate

moving pattern of facts and relationships. The marshalling of chapters, the balancing of sentences, the neat turning of a phrase are regarded as the scholar's hallmark, and little respect is paid to compacter methods of making a statement: the table of figures, the diagrammatic note with its use of symbols (the genealogical tree is a common example), the well chosen illustration, the map. All this is treated as quite separate from writing—if anything, as a branch of typography. Indeed, it is quite often regarded as the publisher's affair.

It is, of course, possible to lapse into the opposite kind of clumsiness, using elaborate pseudo-mathematical notation to say something that is either self-evident or better exposed in a few simple words. There is a certain danger that computer work may lure writers in this direction, that the terminology of the technicians may infect all communications with, by, or about such machines. None the less, it does seem that machine searching and machine storage of facts is likely in the long run to discourage all inefficiencies of communication (just as machine translation must necessarily expose all meaningless rhetoric), and operate to break up the smooth narrative with its artificial links and antitheses into a more disjointed but multidimensional web of different elements, where graphic, numerical, and verbal items will each play a part. Guided by the index, which will need to be much more efficient than it often is at present, the machine will then be able to pick out any item selected. Reading through the book itself, the student will no longer follow a single narrative-cum-explanatory thread but will find a number of relationships suggested simultaneously by symbols or by physical relationship on the page.

There are precedents for this type of condensed, non-linear arrangement which might already be worth re-studying to see whether they will meet the new demands.

There is the layout of advertisements and illustrated magazines, which combine words with visual elements in a quite conscious and purposeful two-dimensional order.

There is the free verse of the poets, which is often akin to tabulation, and the very original verbal-cum-typographic experiments of the Futurists and Dadaists and odd individuals like Apollinaire.

There is the use made of symbols and letters in certain paintings: Klee's black arrows, for instance, which undoubtedly have a "literary" message.

There are instructional textbooks and military orders, where life and death may often depend on logical arrangement and precision of meaning and tables or diagrams alternate with a severely economical (!) prose.

There are Le Corbusier's famous manifesto-like backboard diagrams which are the real kernel of his books.

There are the experiments already going on in the United States for the preparation of printed textbooks for "programmed learning."

Put all these together and examine them, and we might well see the beginnings of a literary technique well suited to the complexity of our age. At least, they would show that such a technique can provide stimulating problems for the writer and, far from being "mechanical," can have its own strong aesthetic appeal.

The future lies with the scholar who can do more than simply manage linear prose, and that is why it is a pity when criticism isolates the text of any work of nonfiction from its illustration, indexing and general ease of handling. For a book of this sort is a whole in which there should be no hierarchy of elements; all must combine to make it a more or less effective and beautiful tool. Once we realize this, it is possible that certain patterns may be distinguishable in the new texture and become familiar just as words and stock phrases now do.

In this way the writer can come to manipulate larger and larger conceptions, and his own horizons will accordingly widen. He will not only have to waste less time on arranging his material but will be able to cut out much of the effort now spent in merely collating elements from different sources and introducing them into the one-way stream of the book; a process that is at present too often like a cross between abstracting and translation.

He will get a chance to think.

L'ENVOI

Any philosophical system is at best a set of internally consistent statements about *a* universe made by *a* particular philosopher at *a* particular period in his intellectual development. It should pretend to nothing more. A philosophical system may well tell us more about *the* philosopher than it does about *the* universe it purports to describe. To that extent, if nothing more, this is a philosophical paper.

I have tried to distinguish in this paper between my personal opinions, based on my own experiences, association nets and classifications on the one hand, and statements I believe to be operationally verifiable and/or statistically valid on the other. The professional philosopher, far better than I at classifying what people have written about the problems discussed in this paper, will find my outlook a curious blend of nominalism, linguistic philosophy, logical empiricism (perhaps even neo-positivism), epistemological solipsism and, I am afraid, the vitalism natural to one trained in the biological sciences.

I have found the concepts outlined in this paper useful in understanding what I, as a conscious entity, am trying to do when I think and try to express my thoughts in words. I claim no eternal verities. I make only the claim of the pragmatist: "At least for me, sometimes, it works."

REFERENCES AND MARGINALIA

1. I am indebted to Jerome Rothstein for the concept of thermodynamics as the casting out of demons. Dr. Rothstein should not be held responsible for any extrapolations I have made of this idea.

2. As cited by H. P. Lovecraft. I have had some difficulty in furnishing proper bibliographic citation, since the only surviving holographic copy is rumored to be located somewhere in Transylvania.

3. The *Third* edition only seems appropriate for this footnote number.

4. The text quoted is that of the Nonesuch Swift, from the Heritage Press 1940 edition of *Gulliver's Travels. An Account of the Four Voyages into Several Remote Nations of the World.*

5. The material which follows is a development of a concept originally and most inadequately set forth in a paper of mine, "Current Trends in Documentation Research," given at a Federation meeting symposium on biomedical documentation in Atlantic City, April 1963, in *Federation Proceedings*, 22: 988—992, July-August, 1963.

6. This delicious phrase is quoted from *Words and Things*, by Roger Brown, Free Press, Glencoe, Ill., 1958. This title should not be confused with that of *Words and Things*, by Ernest Gellner, Beacon Press, Boston, 1959 nor, for that matter, with "Words and Things," in M. W. Thistle's Keynote Address in 1962 at this Institute.

7. My thinking in this area, including the specific use of the word, "tokens" has been rather heavily influenced by a paper by Richard Reiss, "An Abstract Machine Based on Classical Association Psychology," produced under contract with the Air Force Office of Scientific Research and presented at the AFIPS Spring Joint Computer Conference, San Francisco, May 1962. Proceedings available from National Press, Palo Alto, California. It is a curious psychological point that most of us who are considering psychological problems without formal training in the field apparently do so by re-inventing association psychology.

8. Aldous Huxley in "Knowledge and Understanding." Reprinted in *Collected Essays*, Bantam, N. Y., 1960.

9. See, for example, F. S. C. Northrop's introduction to *Physics and Philosophy—The Revolution in Modern Science*, by Werner Heisenberg. Harper, N. Y., 1956.

10. Adam Schaff *Introduction to Semantics*. Pergamon *cum* Macmillan, N. Y., 1962.

11. K. Marx, *Capital*, Vol. I, Foreign Languages Publishing House, Moscow, 1954.

12. S. L. Rubinstein, *Principles of General Psychology*. Moscow, 1946.

13. Uriel Weinreich in "Problems in Lexicography," *International Journal of American Linguistics* 28: 25-43, April 1962.

14. Richard Robinson, *Definition*, Oxford, London, 2nd printing 1954. Quoted via 13, above.

15. H. W. Fowler, *A Dictionary of Modern English Usage*, Oxford, London, 1947.

16. As edited by John Melton in: J. W. Perry and Allen Kent: *Tools for Machine Literature Searching*. Interscience, N. Y., 1958.

17. "Report on the Colloquium on Semantic

Problems in Language," held by the Cambridge Language Research Unit at King's College, Cambridge, England, Sept. 9-10, 1961, under Contract AF 61(052)-545 with the Air Force Office of Scientific Research.

18. Used by T. H. White in a description of Merlin's study in *The Once and Future King*, Dell, N. Y., 1960.

19. The figure is taken from the *1960 World Almanac*. I have not subjected the figure to computational nor experimental verification.

20. Perhaps most recently quoted in Gary Jennings, "Pidgin: no laughing matter," *Harper's*, July 1963.

21. Gibbon, in writing of the various religions, of which the Roman Empire had almost as many as Southern California, said something like: "The philosophers find them equally false; the people find them equally true, and the magistrates equally useful." No attempt should be made to find a 1:1 correspondence between these classes and those mentioned in my text.

22. Aldous Huxley in "The Oddest Science." Citation as in (8), *supra*.

23. W. S. Churchhill in *Roving Commission: My Early Life*, 1930.

24. Mortimer Taube, in a propadeutic private communication, 30 October 1962.

25. The discussion which follows is based in part on reports written by Dr. Herbert Bohnert under Contract AF 49(638)-1198 with the Air Force Office of Scientific Research. At present, Dr. Bohnert has developed LOGOS IV, a symbolic language which provides natural syntactic settings for nouns, adjectives and verbs in the neutral present tense only; in affording full flexibility in the use of relative pronouns (e.g., the grammar encompasses such constructions as "that to which," and in introducing third person singular pronouns and definite composites, such as "the tree."

26. As set forth by Silvio Ceccato in "Linguistic Analysis and Programming for Mechanical Translation; Mechanical Translation and Thought." RADC-TR-60-18, under Contract AF 61 (052)-212, monitored by Rome Air Development Center, U. S. Air Force, Gordon & Breach, N. Y., 1962.

27. Ernst V. Glasersfeld, "Analysis of M aning in Terms of Operations," Rapport CETIS No. 24, EURATOM, C. C. R., Ispra, Italy.

28. In *Tomlinson*, "Ballads and Barrack-Room Ballads."

29. I am indebted to Gordon Peterson of the University of Michigan for reminding me of the importance of the concept of an "array" of signals.

30. C. C. Bombaugh, *Oddities and Curiosities of Words and Literature*, edited and annotated by Martin Gardner, Dover, N. Y., 1961.

31. From a footnote in *The Annotated Alice* by Lewis Carroll, with introduction and notes by Martin Gardner, Potter, N. Y., 1960.

32. Quoted in Anthony G. Oettinger, *Automatic Language Translation. Lexical and Technical Aspects, With Particular Reference to Russian*, Harvard, 1960.

33. "Freeing the Mind—Articles and Letters from the Times Literary Supplement During March -June 1962." Available for 3/6 (approximately 50¢) from the Times Publishing Co., Ltd., Printing House Square, London, EC4, England.

ABOUT DR. HAROLD WOOSTER

Dr. Harold Wooster is the Director of the Information Sciences Directorate of the Air Force Office of Scientific Research. This Directorate has the responsibility for the operation of two Divisions: the Information Research Division and the Technical Information Division. Through the Information Research Division, a multi-disciplinary basic research program is managed in the advanced non-numerical uses of computers in such areas as pattern recognition; lexical processing (including data or information processing, storage and retrieval, and translation); encoding for communication and control; and decision making. The Technical Information Division is the technical library organ for AFOSR and the Office of Aerospace Research.

Dr. Wooster received an A.B. degree in chemistry, *magna cum laude*, from Syracuse University in 1939. He received his M.S. in 1941 and his Ph.D. in 1943, for research in clinical endocrinology, from the University of Wisconsin. During the war years, he worked for the National Defense Research Committee, OSRD, at the Toxicity Laboratory, University of Chicago, in classified research on novel chemical warfare agents. In 1946, he went to the Pepper Laboratory of Clinical Medicine, University of Pennsylvania, under an Office of Naval Research contract.

Dr. Wooster joined the Mellon Institute, Pittsburgh, in 1947, as Senior Fellow on a Food Varieties Fellowship. He combined laboratory research in nutrition and food biochemistry with writing and editing in nutrition. He edited the quarterly journal, *Nutritional Observatory*, and produced the standard reference work and teaching aid, *Nutritional Data*.

In 1956, Dr. Wooster joined the staff of the Air Force Office of Scientific Research. In addition to managing the principal Air Force basic research program in the information sciences, he has edited four books: *Vistas in Astronautics, Information Storage and Retrieval, Basic Research Resumes—a Survey of Air Force Basic Research*, and the *Air Force Scientific Research Bibliography*.

Dr. Wooster is a member of many professional and honorary societies including Phi Beta Kappa and Sigma Xi. His name is listed in *American Men of Science, Who's Who in the South*, and the *Who's Who Supplement*, and the 4th edition of *Leaders in American Science*.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Directorate of Information Sciences Air Force Office of Scientific Research Arlington, Virginia 22209	UNCLASSIFIED
	2b. GROUP

3. REPORT TITLE

A WEB OF WORDS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Scientific   Interim

5. AUTHOR(S) *(First name, middle initial, last name)*

Harold Wooster

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
July 1963	44	33

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
In-house	
b. PROJECT NO 9769	
c. 61445014	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d. 681304	AFOSR 67-2809

10. DISTRIBUTION STATEMENT

1 - Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
PROCEEDINGS Presented at 6th Annual Institute in Technical & Industrial Communications, Ft Collins, Colorado, 8-12 July 1963	Air Force Office of Scientific Research(SRI) 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

The author attempts to apply current notions in linguistics and philosophy to the problems of technical writing. Major sections are: the frame of reference; words and word-books; words plus words make sentences; sentences plus sentences make prose. Typical headings include: Gulliver the gullible, or linguistics in Lagado; the definer defined; Aristotle was class conscious; what your dictionary won't tell you; must we write pidgin?; the mystique of sentencehood; it's art, but is it pretty?; a sentence is a discrete fragment of a correlation network of "thoughts" expressed as a sequential linear array of symbols. The author concludes: "I have found the concepts outlined in this paper useful in understanding what I, as a conscious entity, am trying to do when I think and try to express my thoughts in words. I claim no eternal verities. I make only the claim of the pragmatist:' At least for me, sometimes, it works.'"

DD FORM 1473
1 NOV 65

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT.
Dictionaries						
Thesauri						
Technical writing						
Linguistics						
Demonology						
Semantics						
Syntax						